

ТЕХНОЛОГИЧЕСКИЕ АСПЕКТЫ СОЗДАНИЯ ЭЛЕКТРОННЫХ РЕСУРСОВ ЦЕНТРАЛЬНОЙ НАУЧНОЙ БИБЛИОТЕКИ МОН РК

В условиях формирования современного информационного общества чрезвычайно важным инструментом устойчивого экономического и социального развития является обеспечение публичного доступа к социально значимой информации, в первую очередь научного, образовательного и культурного характера. Стремительное развитие технологий хранения и передачи информации вынуждают библиотеки искать новые пути для сохранения статуса библиотеки как поставщика информационных ресурсов. Переход к новым технологиям открывает новые возможности как для поиска и предоставления информации, так и при выполнении традиционных форм обслуживания. Центральная научная библиотека МОН РК уделяет значительное внимание развитию современных технологий библиотечного дела, выполняя работы по следующим направлениям:

внедрение автоматизированной информационной библиотечной системы “МАРК-SQL” НПО «Информ-система»;

создание баз данных полнотекстовых копий документов;

развитие интернет сайта ЦНБ МОН РК <http://www.library.kz>.

АИБС “МАРК-SQL” (1) является интегрированной информационной системой, предназначенной для автоматизации библиотечного дела. Архитектура системы клиент-сервер, доступ к базам данных обеспечиваются по стандарту ODBC (OLE DB). АИБС “МАРК-SQL” полноценно поддерживает стандарты библиотечной автома-

тизации MARC, включая USMARC и RUSMARC. АИБС “МАРК-SQL” поддерживает собственный макроязык скриптового типа и включает встроенный интерпретатор выполнения скриптов. Одной из особенностей АИБС “МАРК-SQL” является режим “Макрообъекты”, позволяющий описывать разнообразные виды электронных ресурсов, таких, как полнотекстовая информация, мультимедиа, обучающие системы, базы данных и т. д., кроме того, эти ресурсы могут быть добавлены к определенным библиографическим записям. В процессе работы с библиографическими записями пользователь может получить доступ к электронным ресурсам, приписанным выбранным библиографическим записям. В зависимости от типа макрообъекта для его просмотра будет вызвана определенная программа, например редакторы текстов, графики, программа прослушивания, просмотра аудио-видеоинформации. Библиографическому описанию можно приписывать любое количество макрообъектов.

Возможности режима “Макрообъекты” использованы в ЦНБ МОН РК для создания полнотекстовых баз данных. Основной целью создания полнотекстовых баз данных является предоставление пользователям библиотеки сервисов оперативного поиска и доступа к электронным материалам, включающим копии документов в формате PDF. Технология создания полнотекстовой базы данных включает несколько этапов. На первом этапе создается библиографическая база данных, далее к библиографическим записям

документов электронного каталога добавляются их электронные ресурсы по следующей схеме:

1) настройка шаблонов базы данных с использованием технологии макрообъектов;

2) идентификация электронных ресурсов с библиографическими записями;

3) составление поисковых словарей.

В результате разработки технологии по формированию полнотекстовых баз данных в ЦНБ МОН РК созданы базы данных научных статей электронных изданий (архивов) «Русский курьер», «Tiborder» Германия, издательства «Elsevier» и др.

Одной из первоочередных задач информационного обслуживания научной деятельности является обеспечение полноценного, в том числе удаленного, доступа пользователей к электронным информационным ресурсам. Для обеспечения удаленного доступа к ресурсам ЦНБ МОН РК создан Интернет-сайт <http://www.library.kz>, использующий технологии PHP+MySQL и предоставляющий доступ к электронному каталогу, новым поступлениям и другим материалам. Дальнейшее развитие Интернет-сайта ЦНБ МОН РК предполагает реализацию международного стандарта ANSI/NISO Z39.50 (ISO 23950), определяющего протокол поиска и извлечения информации в распределенной неоднородной среде и являющегося основным стандартом для работы с базами данных библиографических каталогов. Для полноценной работы удаленных пользователей с библиографическими данными через Интернет необходимо внедрение программного обеспечения шлюза HTTP-Z39.50.

Большинство запросов пользователей имеют отраслевой характер, поскольку он является основой организации научной деятельности и системы обучения в средней и высшей школе. В связи с этим актуальна задача построения базы данных электронного каталога, ориентированных на отрасли науки. Процесс создания отраслевых баз данных электронного каталога включает научную разработку классификационных схем понятий и выявление терминологического фонда из представительного фонда информационных документов. Классификационные схемы понятий необходимы при обработке поисковых запросов для максимальной детализации области поиска документов. Текст поискового запроса уточняется и дополняется с использованием тождественных трансформаций слов запроса, добавлением сино-

нимов и близких по смыслу слов. Определение близости по смыслу слов является контекстно-зависимым, для создания классификационных схем понятий используются отраслевые тезаурусы, энциклопедические, терминологические и толковые словари, таблицы универсальной десятичной (децимальной) классификации (УДК), тематические рубрикаторы, библиотечно-библиографические классификации (ББК), государственные стандарты и другие источники.

В Центральной научной библиотеке МОН РК сформированы 35 научных библиографических баз данных электронного каталога по основным отраслям науки.

Одной из важных задач обработки научных публикаций является анализ цитируемых документов, включающий задачу построения индекса цитирования. Цитирование документов можно описать в виде ориентированного графа цитирования, в котором возможны циклы. Вершинами графа являются документы, ориентированные ребра отмечают цитаты. Простейший цикл возникает для двух документов А и В, если А ссылается на В и В ссылается на А. При наличии реляционной базы данных публикаций возникает задача построения графа цитирования, которая сводится к задаче транзитивного замыкания отношений. Рассмотрим отношение R(article, citation), в котором article является идентификатором документа X, citation является идентификатором непосредственно цитируемого из X документа. Для нахождения непосредственных цитат документа А достаточно выполнить селекцию отношения R, для нахождения статей, цитируемых из документов цитируемых из А, достаточно выполнить соединение R с R, таким образом для нахождения документов, косвенно цитируемых из А с глубиной цитирования N, необходимо выполнить соединение N-1 отношений R. Задача построения графа цитирования в виде отношения X(article, citation), в котором citation включает косвенно цитируемые документы, может быть описана на языке DATALOG в следующем виде (X есть транзитивное замыкание отношения R):

$$X(a, b) := R(a, b)$$

$$X(a, b) := X(a, c) R(c, b)$$

В современном стандарте языка программирования SQL предложена конструкция

RECURSIVE WITH, позволяющая решать задачи транзитивного замыкания отношения:

```
WITH RECURSIVE NOCYCLE X(article,
citation) AS
SELECT article, citation FROM R
UNION ALL
SELECT X.article, R.citation
FROM X, R
WHERE X.citation = R.article
SELECT * FROM X;
```

В данном запросе отношение X определяется рекурсивно с использованием UNION ALL, NOCYCLE необходимо применять для исключения циклов. X вычисляется рекурсивно до тех пор, пока соединение X и R добавляет в X новые строки. К сожалению, в настоящее время не все существующие СУБД поддерживают конструкцию RECURSIVE WITH, она реализована в IBM DB2 и Microsoft SQL Server 2005. В СУБД Oracle, начиная с ранних версий, существует расширение

SQL для выполнения иерархических запросов START WITH / CONNECT BY, но конструкция RECURSIVE WITH является более универсальной и решает более широкий класс задач.

ЛИТЕРАТУРА

1. <http://www.informsystema.ru>
2. Шрайберг Я.Л. Основные положения и принципы разработки автоматизированных библиотечно-информационных систем и сетей. М., 2000. 130 с.
3. Земсков А.И., Шрайберг Я.Л. Электронные библиотеки. М., 2003. 352 с.
4. Гончаров М.В., Колосов К.А. Практическая реализация библиотечного Интернет-комплекса. М., 2005. 192 с.
5. Tropashko V. SQL Design Patterns: The Expert Guide to SQL Programming. Rampant TechPress, 2007. 264 p.

УДК 004:027

¹ Центральная научная библиотека
МОН РК;

² Центр физико-математических
исследований МОН РК, г. Алматы Поступила 3.10.06г.